



Rapport de projet d'application

Projet : Système d'authentification basé sur la
reconnaissance faciale

Ingé 3 Majeur Intelligence Artificielle S1 2022

Elaboré par :

Clément Cronier, Théophile Chêne et Adrien Tirlemont élèves à l'ESME-Sudria Encadrant :

Madjid Maidi

➤ Résumé :

Nous avons, au cours de ces dernières semaines pu commencer à répondre à une problématique précise : l'authentification basée sur la reconnaissance faciale.

Tout au long du début projet, nous sommes passés par différentes phases de réflexion. Nous avons commencé, par effectuer une recherche en ce qui concerne les systèmes préexistants. Nous avons alors compris que pour constituer un système de reconnaissance faciale, il faudrait décomposer notre problématique en deux parties distinctes. Tout d'abord la détection de visage puis ensuite la reconnaissance de ce visage.

Pour ce qui est du Deep learning, il nous est apparu comme une évidence que pour apprendre à un modèle comment reconnaître des visages, il nous faudrait une base de données suffisamment conséquente. Il a donc fallu, que nous trouvions une base de données volumineuse et diversifiée pour répondre au mieux à notre problème. Nous verrons dans une prochaine partie sur quels critères nous nous sommes basés afin d'établir ce choix et quel a finalement été notre choix.

De plus, une fois notre base de données prête, nous avons commencé à réfléchir à un modèle de réseau de neurones qui serait le plus optimal pour la reconnaissance de visage. Il a fallu étudier quels modèles existaient déjà, quels étaient leurs performances, sur quel type de réseaux de neurones nous devrions nous focaliser... Nous verrons également dans une prochaine partie que nous avons suivi deux principaux axes de développement en parallèle afin d'étudier quelle piste aboutirait aux meilleurs résultats.

Nous tenons à remercier notre tuteur M.Maidi qui, tout au long du début de notre projet, nous a aidé à comprendre les problématiques de notre sujet. Il nous a également permis d'aboutir à une première solution fonctionnelle grâce à ses conseils et au temps qu'il a investi afin que nous puissions réussir au mieux le lancement de notre projet.

Table des matières :

I. Introduction

II. Revue de la documentation

III. Méthodologie de résolution

IV. Déploiement de la solution

V. Résultats obtenus et analyse

VI. Conclusion et perspectives

VII. Références bibliographiques

I. Introduction

Au cours de notre 5ème année à L'ESME Sudria, un projet a été assigné à chaque groupe d'élèves. Pour ce qui est de notre groupe, le sujet porte sur le Deep Learning, plus précisément l'authentification en utilisant la reconnaissance faciale.

Nous avons, par notre année précédente à l'ESME des notions concernant ce domaine grâce à notre projet de 4ème année basé lui aussi sur le Deep Learning. Il s'agissait finalement de reproduire le fonctionnement du cerveau humain et d'appliquer cette technique à notre modèle. Notre objectif principal, était de mettre en place un algorithme permettant une fois le visage détecté de le classer.

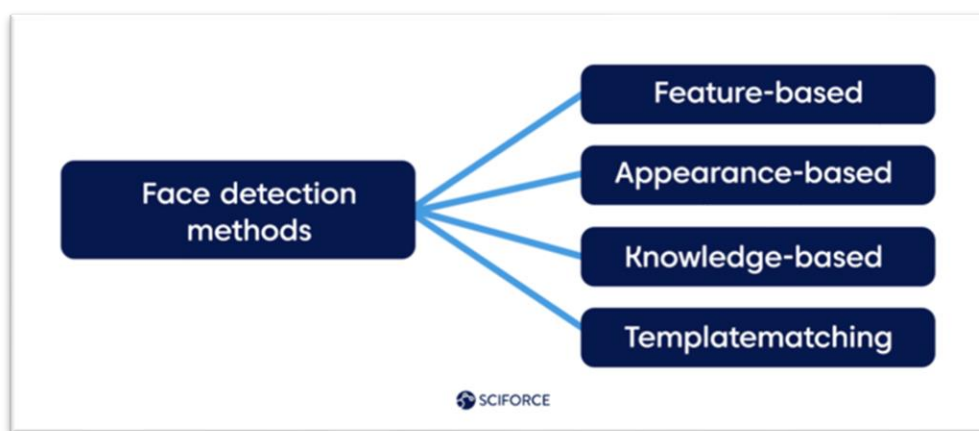
Ainsi, nous avons d'abord imaginé que nous pourrions classer les visages en se basant sur les points d'intérêts de ceux-ci. C'est au cours d'une discussion avec notre tuteur que nous sommes arrivés à la conclusion que nous devrions classer les visages en se basant sur toute l'image et non sur seulement quelques points d'intérêts. Cependant, afin de mener au mieux notre projet nous continuerons à mener en parallèle la solution basée sur l'extraction des points du visage pour leur classification. Une fois, nos recherches menées à bien, nous avons pu mettre en place un plan d'action de façon à structurer notre projet et le diviser en étapes afin de répondre à la problématique le plus efficacement possible.

Finalement, nous verrons premièrement sur quelles études nous nous sommes basés pour bien comprendre le fonctionnement de la reconnaissance faciale et comment nous avons choisi le dataset qui nous servirait à entraîner notre modèle. Ensuite, nous développerons notre idée en passant en revue chacune des étapes clé de l'autre projet. Puis, nous étudierons la solution que nous avons mise en place afin de répondre à la problématique et à quels résultats nous sommes arrivés. Enfin, nous conclurons en interprétant les résultats auxquels nous avons abouti, puis, afin de pousser notre réflexion plus loin, en étudiant les potentiels axes d'amélioration de notre projet.

II. Revue de la documentation

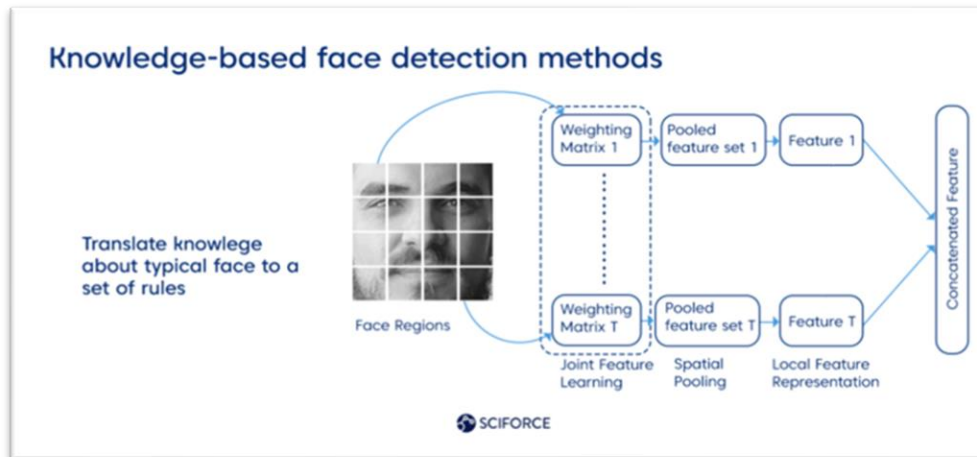
Tout d'abord, afin d'aborder ce projet, il nous a fallu étudier en détails certains concepts afin de commencer le projet de la meilleure façon. Nous avons recherché les différentes méthodes pour la détection de visage.

Etat de l'art de la détection de visage :



Knowledge-based face detection

Cette méthode repose sur l'ensemble des règles élaborées par l'homme selon nos connaissances. Nous savons qu'un visage doit avoir un nez, des yeux et une bouche à certaines distances et positions les uns par rapport aux autres. Le problème avec cette méthode est de construire un ensemble de règles approprié. Si les règles sont trop générales ou trop détaillées, le système se retrouve avec de nombreux faux positifs. Cependant, cela ne fonctionne pas pour toutes les couleurs de peau et dépend des conditions d'éclairage qui peuvent modifier la teinte exacte de la peau d'une personne sur la photo.

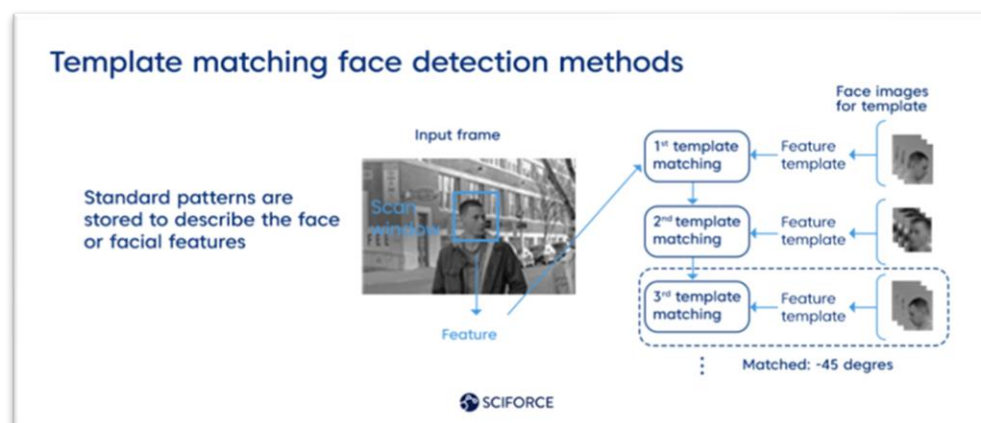


Avantages : facile à comprendre.

Inconvénients : Cela ne fonctionne pas pour toutes les couleurs de peau et dépend des conditions d'éclairage qui peuvent modifier la teinte exacte de la peau d'une personne sur la photo.

Template matching

Le procédé de Template matching utilise des modèles de visage prédéfinis ou paramétrés pour localiser ou détecter les visages par la corrélation entre les modèles prédéfinis ou déformables et les images d'entrée. Le modèle de visage peut être construit par arêtes en utilisant la méthode de détection des arêtes.



Une variante de cette approche est la « controlled background technique ». Si vous avez la chance d'avoir une image de face frontale et un arrière-plan uni, vous pouvez supprimer l'arrière-plan en se basant sur les limites du visage.

Pour cette approche, le logiciel dispose de plusieurs classificateurs pour détecter différents types de visages de face et certains pour les visages de profil, tels que des détecteurs d'yeux, de nez, de bouche et, dans certains cas, même de tout le corps.

Avantages : facile à implémenter.

Inconvénients : C'est de manière générale inapproprié pour la détection de visage.

Feature-based face detection

The feature-based method extrait les caractéristiques structurelles du visage. Il est formé en tant que classificateur, puis utilisé pour différencier les régions faciales et non faciales. Un exemple de cette méthode est la détection de visage basée sur la couleur qui scanne des images ou des vidéos colorées pour les zones avec une couleur de peau typique, puis recherche des segments de visage.

Haar Feature Selection s'appuie sur des propriétés similaires des visages humains pour former des correspondances à partir des traits du visage : emplacement et taille de l'œil, de la bouche, de l'arête du nez et des gradients orientés d'intensités de pixels. Il y a 38 couches de classificateurs en cascade pour obtenir le nombre total de 6061 caractéristiques de chaque face frontale

Avantages : plusieurs modèles pré-entraînés existants, facile à implémenter, bonne accuracy, rapidité d'exécution.

Inconvénients : enclin aux faux positifs/négatifs, régions parfois mal classées ou manquées.

Histogram of Oriented Gradients (HOG) est un extracteur de caractéristiques pour la détection d'objets. Les traits extraits sont la distribution (histogrammes) des directions des gradients (gradients orientés) de l'image.

Gradients ont généralement de grands bords et coins arrondis et nous permettent de détecter ces régions. Au lieu de considérer les intensités de pixel, ils comptent les occurrences de vecteurs de gradient pour représenter la direction de la lumière afin de localiser les segments d'image. La méthode utilise une normalisation de contraste locale superposée pour améliorer la précision.

Avantages : Pas difficile à mettre en place.

Inconvénients : Reconnaît les visages de derrière. Il faudra beaucoup de temps pour l'exécution. Il y aura beaucoup d'informations bruyantes (arrière-plan, flou, éclairs et changements de rotation) que nous ne souhaitons pas considérer comme importantes.

Appearance-based face detection

La méthode « appearance-based » la plus avancée dépend d'un ensemble d'images de visage d'entraînement délégué pour découvrir les modèles de visage. Il s'appuie sur l'apprentissage automatique et l'analyse statistique pour trouver les caractéristiques pertinentes des images de visage et en extraire des caractéristiques.

Avantages : facile à mettre en œuvre, sa robustesse face aux variations de pose, et d'expressions faciales.

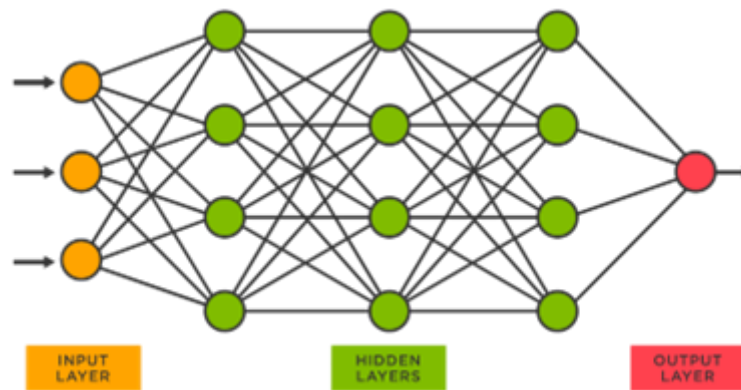
Inconvénients : Couteuse en temps de calcul. Couteuse en espace mémoire. Rend de mauvais résultats quand le nombre d'images d'apprentissage est grand.

Finalement, concernant la détection de visage, nous choisirons de par ses nombreux avantages, la méthode **Haar Feature Selection**, celle-ci ayant une bonne rapidité d'exécution, une bonne accuracy et possédant de nombreux modèles préexistants.

Ensuite, nous nous sommes intéressés aux meilleures méthodes pour la reconnaissance de visage. Parmi les meilleures méthodes, se trouvent les réseaux de neurones convolutifs (CNN). C'est ce type de réseaux que nous avons décidé d'étudier pour pouvoir mener à bien la reconnaissance des visages détectés.

Les réseaux de neurones convolutifs, appartiennent à une sous-catégorie des réseaux de neurones. Ils sont spécialement conçus pour traiter des images en entrée. Leur architecture est plus spécifique, dans la mesure où elle est composée de deux blocs principaux. Le premier bloc fait la particularité de ce type de réseaux de neurones puisqu'il fonctionne comme un extracteur de caractéristiques. La première couche filtre l'image avec plusieurs noyaux de

convolution. Le résultat de ce premier filtrage est ensuite normalisé et redimensionné. Le second bloc, retrouvé dans tous les réseaux de neurones, est utilisé pour la classification.



Réseau de neurones convolutif

Enfin, concernant le choix de notre dataset, nous avons étudié différents datasets. Dans le cadre de notre projet, il nous fallait des images se rapprochant le plus possible d'un cas d'usage réel. Les images doivent être diversifiées, avoir du contraste, des éclairages différents, des conditions réelles finalement.

Nous avons choisi le Face Recognition dataset, une base de données de photographies de visages conçue pour étudier le problème de la reconnaissance faciale sans contrainte. L'ensemble de données contient plus de 4360 images de visages. Il est également composé de 43 classes. Ce dataset possède un nombre d'images considérable avec justement, des images très variées et sans contraintes. Cela répondant parfaitement aux problématiques temps réels de notre système.

Nous avons éliminé certains datasets car ceux-ci contenaient des images qui ne correspondaient pas à notre situation de temps réel ou n'étaient pas assez diversifiées par exemple, ou si le nombre d'images était trop faible. Un cas concret de dataset mis de côté a été le Labeled Faces In The Wild dataset. Nous avons après nos études du dataset remarqué que certaines classes ne contenaient qu'une image rendant ainsi le dataset inutilisable.

III. Méthodologie de résolution

Comme dit précédemment, pour répondre du mieux possible à notre sujet nous avons dû découper notre problème en de multiples étapes.

En premier lieu, nous renseigner sur le fonctionnement des systèmes de reconnaissance faciale. Ensuite, nous avons réfléchi aux diverses méthodes de détection de visage puis à la reconnaissance du visage. Une fois la bonne compréhension du sujet mise en place, nous nous sommes renseignés sur des potentielles bases de données existantes en discutant avec notre tuteur. Base de données que nous avons finalement pu trouver sur Kaggle. Nous avons passé du temps à la compréhension de la structure de la base de données.

Notre base de données était divisée en deux avec un certain nombre d'images destinées à l'apprentissage de notre modèle et le reste des images destinées à la validation.

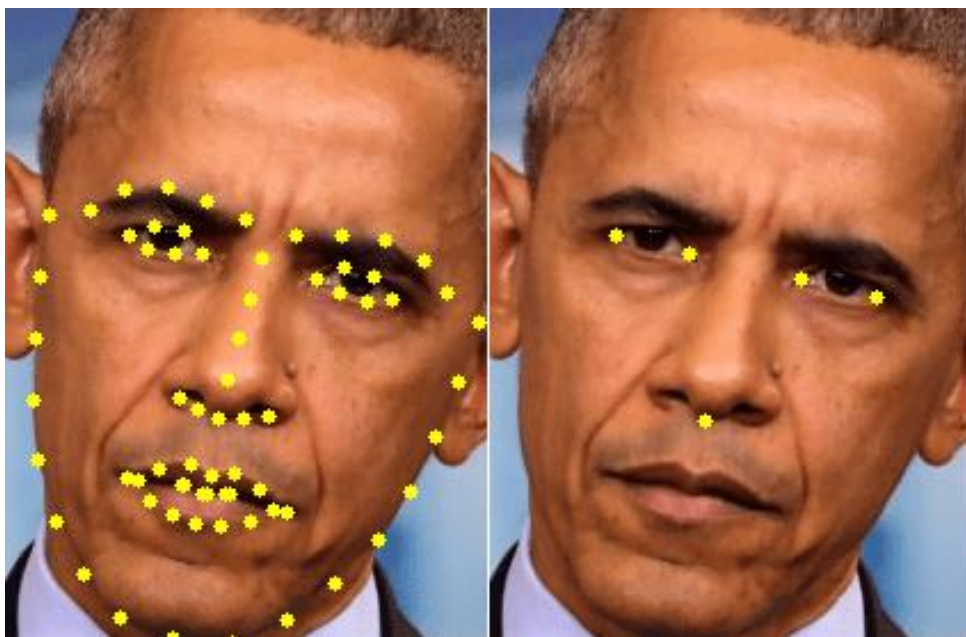
La 2e grande étape a été la mise en place de notre réseau de neurones. En premier lieu nous avons commencé par créer un réseau qui apprendrait à reconnaître les visages en se basant sur les points d'intérêts de ceux-ci. Après échange avec notre tuteur nous avons rapidement vu qu'il serait sûrement plus judicieux de simplement apprendre à notre modèle à reconnaître les visages en lui donnant en entrée toute l'image et non seulement quelques points du visage. Cependant, notre première idée étant intéressante, nous continuerons à explorer cette piste en parallèle de notre réseau de neurones convolutif. Nous utiliserons également un dataset différent dans le cadre de cette approche basée sur les points d'intérêt du visage. Il s'agit du Facial Key Point Detection Dataset qui se compose de 70 000 images. De plus, les images sont très diversifiées et correspondent à des situations en temps réel. L'approche étant différente, nous devons adapter le choix du dataset.

IV. Déploiement de la solution

Pour pouvoir répondre à cette problématique, comme dit précédemment, nous nous sommes renseignés sur le fonctionnement des réseaux de neurones convolutifs. Nous étudierons ici en plus de la méthode de détection de visage, les deux méthodes évoquées précédemment pour la reconnaissance des visages :

1. Méthode de reconnaissance basée sur l'extraction des points d'intérêts du visage :

La méthode de reconnaissance basée sur l'extraction des points d'intérêt se base sur les données biométriques identifiables sur une photo. Ce sont des points d'intérêt comme le centre des iris, la forme du menton, de la bouche etc... Dans cette division du domaine de la reconnaissance faciale, le nombre de marques du visage évolue entre 5 et 68. Evidemment, plus nous avons de marque, plus le système est sécurisé.



Visage avec 68 points

Visage avec 5 points

Pour pouvoir apprendre à notre modèle à tracer les points, il faut décider de ces données d'entrées, de sortie mais surtout de son type. Ici, nous cherchons à lui faire tracer des points. Ce sont des données réelles, donc nous voulons définir une régression. Les données d'entrées seront les images, et les données de sortie seront les 68 points (donc 136 valeurs, pour les coordonnées (x,y)). Comme dit précédemment, le Dataset choisi, met à disposition 70 000 images pour lesquelles nous avons 68 points d'intérêts.

2. Méthode de reconnaissance basée sur l'image entière :

Dans la reconnaissance d'objet basée sur l'apparence, les caractéristiques sont choisies pour être les valeurs d'intensité de pixel dans une image de l'objet. Ces intensités de pixel correspondent directement à l'éclat de la lumière émise par l'objet le long de certains rayons dans l'espace.

Ainsi, ces caractéristiques sont extraites d'une image du visage et utilisées pour créer une représentation numérique, ou "descripteur de visage", qui capture les caractéristiques uniques de ce visage. Le modèle peut ensuite comparer le descripteur de visage, d'un visage inconnu à une base de données de descripteurs de visages connus et déterminer si le visage appartient à une personne enregistrée dans la base de données.

Les modèles basés sur l'apparence sont généralement formés sur un grand ensemble de données d'images de visages qui ont été étiquetés avec l'identité de la personne dans l'image. Le modèle est alors capable d'apprendre les variations d'apparence faciale associées à différentes personnes. Lorsqu'il est présenté avec un visage inconnu, le modèle utilise ces connaissances apprises pour déterminer à quel visage connu le visage inconnu ressemble le plus.

V. Résultats obtenus et analyse

Nous rappelons que la précision (accuracy en anglais) est une métrique de performance qui évalue la capacité d'un modèle de classification à bien prédire à la fois les individus positifs et les individus négatifs.

De la même manière, la perte (loss en anglais) est une métrique qui évalue l'écart entre les prédictions réalisées par le réseau de neurones et les valeurs réelles des observations utilisées pendant l'apprentissage.

Ces deux métriques nous ont été très utiles pour juger des performances de nos modèles. Nous cherchions à maximiser la précision tout en minimisant la perte.

1. Résultats obtenus pour la méthode de reconnaissance basée sur l'extraction des points d'intérêts du visage :

Cette méthode constituant une régression, nous nous baserons essentiellement sur la perte pour pouvoir évaluer ses performances. Cette perte sera définie par l'erreur quadratique moyenne. Nous avons choisi cet estimateur, car nous voulons pénaliser les grandes erreurs pour pouvoir obtenir des points les plus représentatifs possibles de la réalité. Nous avons cherché à la minimiser au cours de nos différents entraînements. Après nos divers entraînements, nous obtenons une erreur quadratique moyenne de 0.0177. Ceci est une erreur très faible. Nous avons ensuite tracé les points d'intérêt sur un de nos visage pour observer s'il tracer correctement les points sur des données inconnues.



Image inconnu du dataset

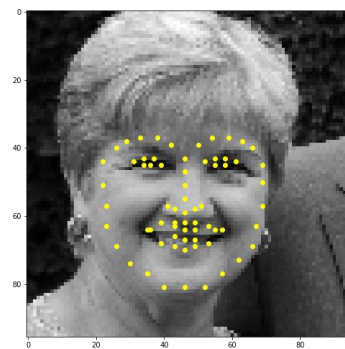


Image de test du Dataset

2. Résultats obtenus pour la méthode de reconnaissance basée sur l'image entière :

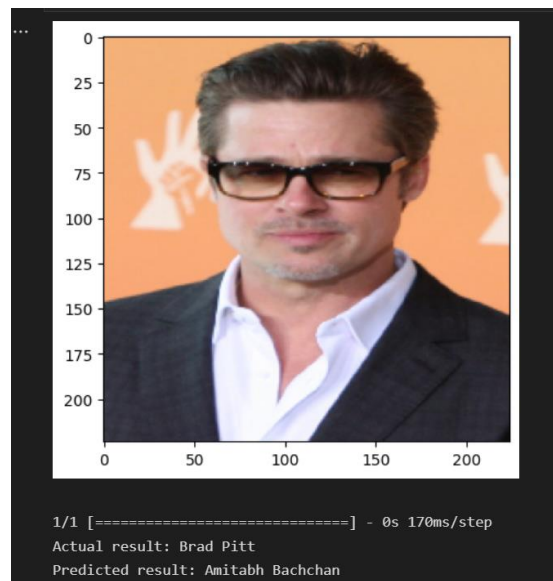


Figure : Exemple d'image utilisé

Pour la méthode basée sur l'image entière, nous avons choisi deux indicateurs de performance comme la perte et la précision catégorique. Nous obtenons ainsi les résultats suivants, une perte de 0.1608, et une précision catégorique de 0.9514. A la vue de ceux-ci, on peut en déduire qu'il y a une beaucoup de mauvaises prédictions. Nous chercherons à améliorer les prédictions. D'autre part pour l'accuracy, le résultat est bon mais on cherchera par la suite à obtenir un résultat se rapprochant de 98-99%, en essayant différents modèles, de manière à obtenir une reconnaissance la plus fiable possible.

VI. Conclusion et perspectives

En conclusion de notre projet, nous avons dû établir un système de reconnaissance faciale capable de détecter puis de classifier le visage. Notre principal problème a été l'utilisation du Deep Learning. Nous avons rencontré et surmonté plusieurs difficultés. Pour résoudre de manière optimale notre problématique, nous nous sommes basées sur deux approches distinctes afin d'étudier quelle serait la meilleure pour la classification. L'une correspondant à un réseau de neurones convolutif apprenant sur des images comprenant l'ensemble du visage et une autre approche basée sur des images contenant les points d'intérêt de ceux-ci. Les deux modèles ont donné des résultats plutôt satisfaisants mais pouvant encore être améliorés.

VII. Références bibliographiques

<https://www.kaggle.com/datasets/prashantarorat/facial-key-point-data>

<https://www.kaggle.com/code/kerimelebiler/face-recognition-with-cnn/data>

<https://medium.com/sciforce/face-detection-explained-state-of-the-art-methods-and-best-tools-f730fca16294>

<https://www.sicara.fr/blog-technique/2019-09-26-face-detectors-dsfd-state-of-the-art-algorithms>

<https://fr.blog.businessdecision.com/tutoriel-deep-learning-le-reseau-neuronal-convolutif-cnn/>